CONGRESS ON NEW RESEARCH IDEAS IN ENGINEERING SCIENCE, TECHNOLOGY, ELECTRICAL & COMPUTER SCIENCE
16 MAY, 2018 SARI - IRAN
کنگره ایده های نوین پژوهشی در علوم مهندسی و تکنولوژی، برق و کامپیوتر
ساری - ایران ۲۶ اردیبهشت ۹۷
شماره مجوز isc وزارت علوم: ۹۷۱۸۰-۱۲۷۰۳
ISC
SETCONG

طبقه بندی تصاویر ثبت شده از راه دور با استفاده از الگوریتم یادگیری عمیق CNN

استاد راهنما : جناب آقای حمید رضا مقسمی

استاد مشاور : سرکار خانم الهام وصلی

صفا منافی بجوشین

دانشجوی کارشناسی ناپیوسته مهندسی تکنولوژی نرم افزار، موسسه آموزش عالی فاران مهر دانش

Safa.manafi@yahoo.com

چکیده

امروزه الگوریتم ها و مدل های مختلف پژوهش های مبتنی بر شبکه عصبی، جای خود را در میان طبقه بندی تصاویر به خوبی باز کرده اند. هدف اصلی این الگوریتم ها این است که در شبکه های مصنوعی، ماشین به شکلی آموزش ببیند که در نهایت تشخیصی نزدیک مغز انسان داشته باشد . از بین انواع شبکه های عصبی، شبکه های عصبی کانالوشن (CNN) معمولا دقت خوبی را در طبقه بندی تصاویر ارائه می کنند . در این مقاله، بهرگیری از یادگیری عمیق در سنجش از راه دور را توسط سه استراتژی متفاوت ارزیابی و آنالیز می کنیم. در بسیاری از برنامه های کاربردی، مخصوصا برنامه های سنجش از راه دور، به علت هزینه های محاسباتی بالا و نیاز به مقادیر بالای داده های برچسب دار، امکان طراحی و آموزش شبکه های عصبی کانالوشن جدید وجود ندارد. آزمایش های این تحقیق با بهره بردن از مجموعه داده سنجش از راه دور و همچنین شبکه های عصبی کانالوشن معروف (fine-tunned) ، صورت می گیرد . نتایج نشان می دهد که شبکه های عصبی کانالوشن به خوبی تنظیم شده ، دارای بهترین عملکرد در بین استراتژی ها می باشند. در حقیقت استفاده از ویژگی های شبکه های عصبی کانالوشن به خوبی تنظیم شده با Linear SVM بهترین نتیجه را می دهد . در حقیقت، با استفاده همزمان از ویژگی های شبکه های عصبی کانالوشن به خوبی تنظیم شده به همراه SVM های خطی تنظیم شده، بهترین نتیجه به دست می آید. هدف اصلی این مقاله ارزیابی استراتژی های مناسب برای بهره برداری بیشتر از توانایی های یادگیری عمیق جهت طبقه بندی صحنه های تصویری ماهواره ای و سنجش از راه دور است.

کلمات کلیدی : یادگیری عمیق، شبکه های عصبی کانالوشن ، به خوبی تنظیم شده ، سنجش از راه دور

۱- مقدمه

امروزه با انفجار اطلاعات و داده روبرو هستیم ، پیدا کردن روشی که بتوان این پایگاه داده های عظیم را پردازش کرده و از آنها اطلاعات مورد نظر را بدست آورد نیز خود یک جنبه مهم در پژوهش ها می باشد. در این مقاله به موضوع پردازش تصاویر بزرگ و از راه دور [۱۴،۱۵،۱۶] با تکیه بر علم هوش مصنوعی و شبکه های عصبی کانالوشن عمیق پرداخته شده است [۳،۴،۵،۶،۷،۸،۹] .

در استراتژی اول (شبکه عصبی کانالوشن کاملا آموزش دیده) یک شبکه یا از قبل وجود دارد یا یک شبکه جدید از صفر آموزش داده می شود بدین صورت که ویژگی های بصری [۱] مشخصی را در مجموعه داده آن قرار می دهیم . اهمیت کار از این جهت می باشد که از طریق آن کنترل کامل معماری و پارامتر ها حاصل می گردد که در نتیجه به شبکه ای کاراتر و موثرتر دست خواهیم یافت .

به دلیل اینکه شبکه ما مستعد خطر overfitting است، این روش مقدار قابل توجهی داده لازم دارد و اگر مجموعه داده ما کوچک باشد این مشکل بزرگتر هم می شود . این ایراد حتی طراحی کامل شبکه و آموزش آن از صفر را برای اکثر مسئله های سنجش از راه دور غیر عملی می کند، به این دلیل که مجموعه داده های بزرگ در این نوع دامنه ها غیر معمول است و همچنین آموزش آنها احتمالا هزینه های زیادی دارد.

در تلاش برای غلبه به این مشکل می توان از داده های کم برای آموزش شبکه استفاده کرد و تکنیک بزرگتر کردن داده (data augmentation) را برای آن به کار برد . ذکر این نکته ضروری است که اجرای این روش برای مجموعه داده های کوچک ، پاسخ مناسبی ارائه نمی دهد.

استراتژی دوم (شبکه عصبی کانالوشن به خوبی تنظیم شده) از یک شبکه از پیش آموزش داده شده استفاده می کند و پارامترهای آن را با استفاده از داده های مورد استفاده برای سنجش از راه دور، پردازش کرده و به تنظیم شبکه می پردازد . معمولا در این روش لایه های اولیه که نقش کلی دارند ، نگه داشته شده و لایه های نهایی با توجه به زمینه ای که برنامه قرار است در آن مورد استفاده قرار بگیرد تعیین می شوند و برای رمزگذاری کردن آن داده ها به کار می روند.

استراتژی سوم (شبکه عصبی کانالوشن از پیش آموزش دیده) به سادگی یک شبکه عصبی کانالوشن از پیش آموزش داده شده را به عنوان استخراج کننده ویژگی به کار می برد، که در آن آخرین لایه‌ی طبقه بندی حذف شده و لایه های قبل آن به عنوان بردار ویژگی داده ورودی ، استفاده می شوند.

می توان چالش هایی که این مدل برای رویارویی با آنها طراحی شده است را به صورت زیر در نظر گرفت :

- مقابله با مساله overfitting در طبقه بندی مجموعه تصاویر بزرگ .
- در نظر گرفتن مدل های قدرتمندتر برای آموزش داده ها با استفاده از مجموعه های آموزشی بسیار بزرگ تر .
- استفاده از GPU های کارآمد برای بالا بردن سرعت عملکرد.

۲- پیشینه مطالعه

تلاش‌های قابل توجهی برای گسترش یک توصیف گر مناسب جهت برنامه های سنجش از راه دور اختصاص داده شده است. اگرچه تعداد زیادی از این توصیف گرها به طور موفقیت آمیز برای پردازش عکس در سنجش از راه دوراستفاده شده‌اند، اما بخاطر اطلاعات غیرقابل دیدن با چشم انسان که در چندین محدوده‌ی دیداری مختلف موجود است و همچنین عدم سازگاری روش های فعلی با صحنه های ثبت شده از راه دور ، جهت ارائه تکنیک های دقیق تر درخواست های زیادی شده است. برای رسیدن به بهترین نتایج ، شبکه های عصبی کانالوشن [۱۲،۱۳] را به کار می گیرند. آنها از یادگیری End-To-End استفاده می‌کنند که به معنی تبدیل داده های خام یا پیکسل به کد می باشد که حتی با وجود اینکه در طول مراحل یادگیری فقط پارامترها یاد گرفته می‌شوند و نه کل معماری عمیق، باز هم این کار مزیت بزرگی در مقایسه با متدهای قبلی محسوب می‌شود [۲]. شبکه های عصبی عمیق کانالوشن چندین ایراد دارند مانند: (۱) هزینه‌ی محاسباتی بالا (۲) مستعد خطر overfitting بودن (۳) تجربی بودن توسعه مدل.

وقتی که با شبکه های عصبی کانالوشن کار می‌کنیم جهت آموزش یک شبکه جدید از صفر ، این دو استراتژی اساس استراتژی کلی را شکل می‌دهند: (۱) آموزش یک شبکه (جدید یا موجود) از ابتدا، به این دلیل بهتر است که ما می توانیم فقط ویژگی های مربوط به زمینه ای که می خواهیم در آن شبکه استفاده کنیم را به عنوان مجموعه داده به آن بدهیم. همچنین این استراتژی کنترل کامل معماری و پارامترها را نیز بدست می آورد، که در نهایت شبکه کاراتر و سریعتر و به اصطلاح robust را خواهیم داشت. در طول سال ها شبکه های عصبی کانالوشن موفق آنهایی بوده اند که با داده های بزرگ آموزش داده شده بودند [۱۰] مانند مجموعه داده های ImageNet ، که از آن برای آموزش چندین معماری مشهور استفاده شده است [۲۱،۲۲،۲۳].

به طور کلی شبکه های عصبی کانالوشن یک ویژگی عجیب و خاص دارند و آن این است که همه‌ی آنها باید لایه‌ی اول را یاد بگیرند که شامل Gabor-filter و تشخیص دهنده‌ی لبه و لکه رنگ می شود. (۲) با توجه به این ویژگی، استراتژی دیگری که می توان با آن از شبکه های عصبی کانالوشن بهره برد، این است که تنظیم دقیق شبکه را برای پارامترهای آن با استفاده از داده های جدید به کار بریم [۱۱]. محققان نشان دادند که انجام تنظیمات دقیق روی یک شبکه عصبی کانالوشن از پیش آموزش داده شده ، روی یک نوع داده مخصوص (داده‌ی هدف) می‌تواند کارایی را بشدت بهبود بخشد. آنها AlexNet را که به وسیله ی Krizhevsky و همکارانش جهت تشخیص تصویر در مقیاس بزرگ ImageNet (ILSVRC) [۱۰] پیشنهاد شده است را به خوبی تنظیم کردند و از نتایج آن برای تقسیمات مفهومی استفاده کردند. Zhao و همکارانش [۱۹] چندین شبکه را دقیق تنظیم کردند که بخوبی موفق به طبقه بندی مجموعه داده های معمولی شدند. [۱۸] همچنین محققان یک شبکه عصبی کانالوشن کاملا آموزش دیده را در مقابل یک شبکه به خوبی تنظیم شده‌ی از آن ارزیابی کردند تا ضعف در سنجش از راه دور آنها را شناسایی کنند . Yue و همکاران [۱۶] روش تنظیم دقیق را برای طبقه بندی عکس های فرابینایی به کار بردند.

براساس ویژگی های فوق الذکر، شبکه های عصبی کانالوشن همچنین می توانند به عنوان استخراج کننده ویژگی در زمینه های دیگر هم بکار گرفته شوند. بطور خاص این ویژگی ها (که معمولا ویژگی های عمیق هستند) با حذف لایه طبقه بندی و در نظر گرفتن خروجی لایه های قبلی بدست می آیند . در بعضی مطالعات اخیر [۱۴،۱۹،۲۰] ، شبکه های عصبی کانالوشن

نشان داده اندکه حتی برای مجموعه داده هایی متفاوت از مجموعه داده ای که برای آن آموزش دیده اند نیز بخوبی عمل کرده اند[۱۹] .

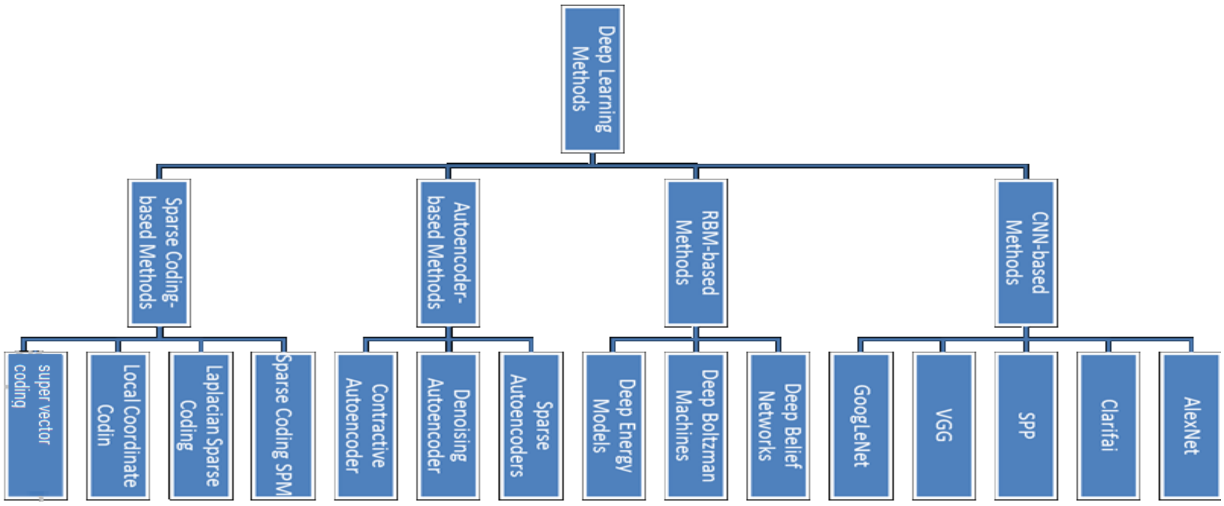
۳- مفاهیم پیش زمینه ای

طی سالهای اخیر, یادگیری عمیق [۱۲،۱۳] در حوزه های مختلف بصورت گسترده مورد مطالعه قرار گرفته است و به همین دلیل تعداد زیادی از روش های مرتبط با آن بوجود آمده است . این روش ها را بر اساس روش پایه ای که از آن مشتق شده اند می توان به ۴ دسته مختلف تقسیم کرد که عبارتند از:

- Convolutional neural networks
- Restricted Boltzmann Machines :RBMS
- Autoencoders
- Sparse Coding

دسته بندی روش های یادگیری عمیق را به همراه کارهای انجام شده در هر یک از این روش ها در شکل ۱ مشاهده می کنید:

.



شکل ۱ : روش های آموزش عمیق در یک نگاه

این بخش به طور کلی برخی مفاهیم شبکه های عصبی کانالوشن و در واقع یک نوع خاص از روش یادگیری عمیق را بیان می‌دارد. این شبکه ها به طور کلی می توانند خروجی هایی که ممکن است برای واحد های بعدی مورد استفاده باشند را با استفاده از داده های ورودی محاسبه کنند.

این نورون ها همگام با هم کار می کنند تا مشکل مشخصی را حل کنند، یادگیری با مثال یک شبکه برای یک درخواست مشخص مثل تشخیص الگو ها یا طبقه بندی داده ها در طول یک پردازش یادگیری ایجاد می شود. همان طور که معرفی شد، شبکه های عصبی کانالوشن در ابتدا برای کار بر روی عکس ها پیشنهاد شدند به این دلیل که آنها سعی در استفاده اهرمی از ویژگی طبیعی ایستا بودن عکس را داشتند . مثلا اطلاعات استخراج شده از یک بخش از عکس می تواند به بخش دیگر نیز اعمال شود. علاوه بر این شبکه های عصبی کانالوشن مزایای مختلف دیگری نیز دارند : (۱) به طور اتوماتیک استخراج کننده‌ی

ویژگی local را یاد بگیرند (۲) در مقابل خطاهای کوچک در ورودی خیلی تغییر نمی کنند (۳) اصول به اشتراک گذاری وزن که به طور شدید تعداد پارامترهای آزاد را کاهش می‌دهد را بکار می‌گیرند و بنابراین قابلیت تعمیم را افزایش می‌دهند.

در ادامه ما بعضی مفاهیم به کار رفته در شبکه های عصبی کانالوشن را بیان می‌کنیم.

۱-۳- واحد های پردازشی

نورون های مصنوعی اصولا واحد های پردازشی هستند که بعضی عملیات محاسباتی را روی چندین متغیر ورودی انجام می‌دهند و معمولا یک خروجی محاسبه شده را در طول تابع فعال سازی تولید می‌کنند . به طور نوعی یک نورون مصنوعی یک بردار وزن $w=(w_1 , w_2 ,..., w_n)$ و بعضی متغیرهای ورودی $X=(x_1 , x_2 ,... , x_n)$ و یک آستانه یا جهت b (bias) را دارا می باشد . از دید ریاضی، بردار های X و W بعد یکسانی دارند . پردازش کامل نورون ممکن است به شکل معادله ۱ بیان شود:

$$z = f\left(\sum_{i}^{N} x_i \times w_i + b\right)$$

(۱)

در حالی که z و x و w وb به ترتیب بیان کننده‌ی خروجی و ورودی و وزن و آستانه می‌باشند. $F(.):R \rightarrow R$ تابع فعال ساز را بیان می دارد . به طور قراردادی، یک تابع غیر خطی برای F(.) درنظر گرفته می شود . توابع مورد استفاده زیادی به عنوان تابع فعال ساز وجود دارد، مانند: Sigmoid ، Herbolic و همچنین تابع اصلاح شده‌ی خطی .

تابع اصلاح شده‌ی خطی ، بیشترین موردی است که در این مقاله استفاده شده است. نورون هایی با این مشخصات وقتی که با دیگر نورون ها مقایسه می‌شوند ، چندین مزیت دارند: (۱) بهتر کار می‌کنند طوری که از اشباع برنامه در حین پردازش یادگیری دوری می‌کنند (۲) موجب عدم تراکم در واحدهای مخفی می‌شوند (۳) مانند توابع Sigmoid و Tanh دچار مشکل ناپدید شدن شیب نمی‌شوند.

واحد پردازشی که از تابع اصلاح گر به عنوان تابع فعال ساز استفاده می کند، واحد فعلی اصلاح شده (ReLU) نامیده می‌شود. اولین گام تابع فعال ساز ReLU در معادله‌ی ۱ نمایش داده شده در حالی که دومین گام در معادله ۲ معرفی می‌شود.

$$a = f(z) = max(0, z) \quad \Leftrightarrow \quad a = \begin{cases} z, if z > 0 \\ 0, otherwise \end{cases}$$

(۲)

۲-۳- اجزا شبکه

در میان انواع مختلف لایه ها، مسئولیت استخراج ویژگی ها از عکس ها با لایه‌ی کانالوشن است. لایه های اول معمولا ویژگی های سطح پایین (مانند لبه ها و خط ها و گوشه ها) را دریافت می کنند در حالی که بقیه لایه ها ویژگی های سطح بالا (مانند قواعد، اشیا و شکل ها) را دریافت می کنند . پردازشی که در این لایه انجام می شود می تواند به دو فاز تقسیم شود: (۱) گام کانالوشن که یک پنجره با اندازه ثابت با یک Stride (گام یا فاصله مشخص بین مرکز پنجره و مرکز عکس) روی عکس اجرا می شود و محدوده مورد نظر را تعریف می کند (۲) مرحله پردازش که از پیکسل موجود در پنجره به عنوان ورودی برای نورون ها استفاده می کند که سر انجام عمل استخراج ویژگی را از آن ناحیه مورد نظر انجام می دهد . اصولا در مرحله ی آخر، درست مانند معادله‌ی ۱ هر پیکسل در وزن متناظر با خودش ضرب می شود و خروجی نورون را تولید می کند، بنابراین

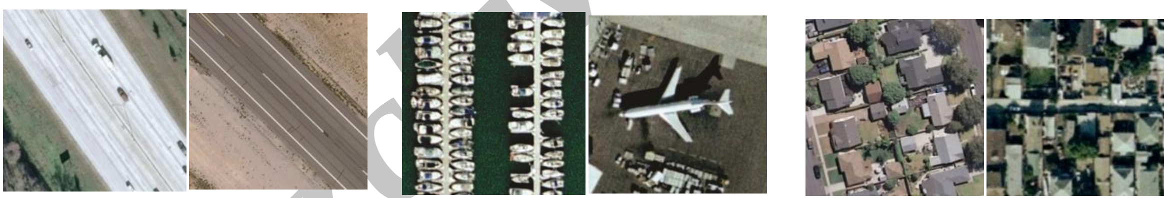
تنها یک خروجی جدید کوچکتر از عکس اصلی را بدست می دهد. اکثر این ویژگی ها شبیه به هم هستند زیرا که هر پنجره می تواند پیکسل های پنجره دیگر را نیز در خود داشته باشد . تغییرات ویژگی ها بوسیله‌ی یک سری از عملیات ها روی ویژگی مشخص یک ناحیه خاص از عکس ایجاد شده است . مشخصا پنجره‌ی با اندازه‌ی ثابت روی ویژگی های استخراج شده با لایه کانالوشن اجرا می‌شود و در هر مرحله عملیات بهینه سازی می شود . روی لایه های pooling دو عملیات وجود دارند . عملیات بیشترین و عملیات میانگین، که بیشترین مقدار و مقدار حد وسط (به ترتیب) را روی ویژگی مورد نظر انتخاب می‌کند. این عملیات تضمین می‌کند که حتی وقتی که ویژگی های عکس تغییرات و چرخش های کوچکی را داشته باشند ، همچنان نتیجه یکسان بماند ، که برای تشخیص و طبقه بندی اشیا ویژگی بسیار مهمی محسوب می شود.

۴- مجموعه داده ها

ما برای ارزیابی بهتر کارایی و مؤثر بودن هر استراتژی ، مجموعه داده هایی با ویژگی‌های تصویری متفاوت را انتخاب کرده‌ایم. اولین مجموعه داده شامل تصاویر هوایی مربوط به زمین با رزولوشن بالا در طیف مرئی می‌باشد . دومین مجموعه داده شامل تصاویر چند سطحی با رزولوشن بالا، با تصاویر جمع‌آوری‌شده از منطقه های مختلف از سراسر دنیا است. سومین مجموعه داده شامل تصاویر چند طیفی با رزولوشن بالا از زمین های قهوه و غیر قهوه می باشد .

۴-۱- مجموعه داده UCMerced

این مجموعه داده که به طور دستی نام گذاری شده و در دسترس عموم قرار دارد از ۲۱۰۰ تصویر هوایی با ۲۵۶*۲۵۶ پیکسل که به طور مساوی به ۲۱ کلاس (سطح) تشکیل شده که که شامل تصاویر کشاورزی، هواپیما، الماس بیس‌بال، ساحل، ساختمان‌ها، بوته‌زار، منطقه انبوه مسکونی، جنگل، آزادراه، زمین گلف، بندرگاه، محل تقاطع راه، منطقه مسکونی نیمه انبوه، پارک خانگی متحرک، روگذر، پارکینگ می باشد که بعضی نمونه سطح‌ها در شکل۲ نشان داده شده‌اند.



شکل ۲ : مجموعه داده UCMerced

۴-۲- مجموعه داده تصاویر گوگل ارث

این مجموعه داده [۱۷] ، شامل ۱۰۰۵ تصویر با رزولوشن فضایی بالا با ۶۰۰*۶۰۰ پیکسل است که به ۱۹ سطح با تقریباً ۵۰ تصویر در هر سطح ، تقسیم‌شده است که از گوگل ارث استخراج شده است ، این مجموعه داده نمونه هایی جمع آوری شده از منطقه های متفاوت از سراسر دنیا را در خود دارد که باعث افزایش گوناگونی آن می‌شود اما چالش‌هایی را نیز به خاطر تغییرات رزولوشن، مقیاس، و چرخش و روشنایی تصاویر ایجاد می‌کند . بعضی از سطوح این مجموعه داده در ۳ نشان داده شده اند.

شکل ۳ : نمونه ای از مجموعه داده تصاویر گوگل ارث

۳-۴- مجموعه داده زمین های قهوه و غیر قهوه

این مجموعه داده [۱۴] از صحنه‌های multi-spectral گرفته شده که به وسیله‌ی حس‌گر SPOT از مزارع قهوه درکشور برزیل تشکیل شده است . تصاویر هر بخش به چندین مربع ۶۴*۶۴ تقسیم شده است ، که ۲۸۷۶ تصویر که به طور مساوی به دو سطح قهوه و غیر قهوه تقسیم شده‌اند. بعضی نمونه های این مجموعه داده در شکل ۴ نشان داده شده‌اند . لازم به ذکر است که این تصاویر از سبز، قرمز و باند نزدیک مادون قرمز تشکیل‌شده است که کاربردی ترین و نمایشی‌ترین مجموعه برای تشخیص نواحی گیاهی می باشد .

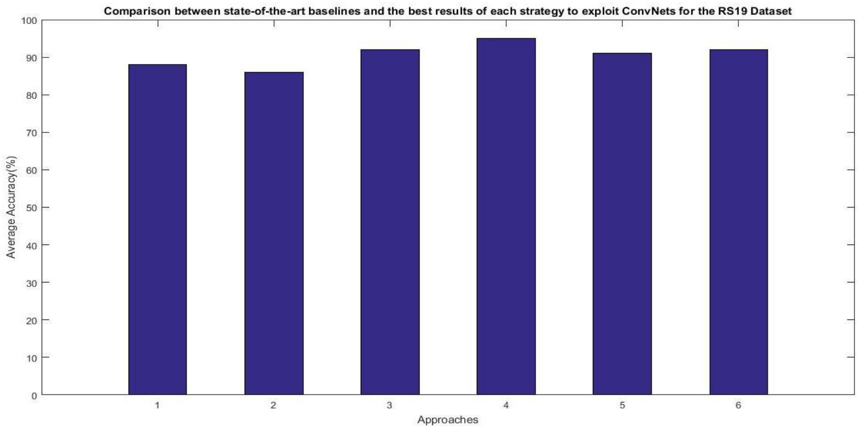
شکل۴: مجموعه داده زمین های قهوه و غیر قهوه

۵- روش اجرای تحقیق

روال کلی کار در این برنامه به این صورت است که، تعدادی تصویر را دریافت می کند، سپس با استفاده از الگوریتم شبکه عصبی کانالوشن یا یادگیری عمیق ، ویژگی ها را از تصاویر استخراج می کند و خروجی را به SVM می دهد. SVM براساس این ویژگی ها روی تصاویر آموزش می دهد و یک مدل برای خودش تولید می کند و سپس محاسبه می کند که تصاویر با چه ویژگی ها در چه دسته هایی قرار گیرند . به طور کلی، تعداد تکرار مسئله به تعداد تصاویر بستگی دارد ، هزار عدد تصویر ۱۰*۱۰ بار باید تکرار شود . بعد ۲۰ درصد از تصاویر به عنوان تست به SVM داده می شود و SVM براساس آموزشی که قبلا دیده اقدام به دسته بندی تصاویر می نماید . دستورات مفیدی در برنامه متلب پیاده سازی گردید و پس از پیاده سازی نتایج زیر حاصل شد:

الگوریتم یادگیری عمیق فرآیند استخراج ویژگی ها از تصاویر را انجام می دهد . الگوریتم SVM قابلیت آن را ندارد که مستقیما روی تصویر کار کند ، به همین دلیل قبل از آن یک استخراج ویژگی انجام می دهیم . به عنوان مثال ، یکسری اعداد و ارقام از تصویر استخراج کرده و به الگوریتم SVM می دهیم که براساس الگوریتم SVM مراحل آموزش ، ارزیابی و درنتیجه طبقه بندی براساس این ویژگی ها انجام پذیرد . در الگوریتم یادگیری عمیق ضمن اینکه هر ویژگی استخراج می شود ، فرآیند آموزش هم صورت می گیرد به طوری که در مرحله اول تصاویر خوانده و سپس Load می گردد و در مرحله دوم ، به

تعداد کلیه تصاویر الگوریتم شبکه های عصبی کانالوشن اجرا می شود سپس کلیه ویژگی ها استخراج می شوند و به الگوریتم SVM داده می شود و در نهایت الگوریتم SVM_CNN اقدام به دسته بندی تصاویر می کند . خروجی برنامه در شکل ۵ نمایش داده شده است.



شکل ۵ : خروجی برنامه

جهت شبیه سازی ابتدا کلیه داده های مربوطه را به سیستم وارد نموده و برای هر پوشه ضمن تعیین یک نام ، برنامه شبیه ساز را از اسامی مربوطه مطلع می سازیم.

۱-۵- ورود مجموعه داده های تصاویر

در این فاز از روش پیشنهادی کلیه دیتاست های معرفی شده به صورت گروهی و مرحله به مرحله به سیستم وارد شده و هر کدام به صورت جداگانه مورد ارزیابی قرار می گیرند. پس از اینکه داده های مربوطه بارگذاری شده و برای هر دیتاست مجموعه تصاویر مربوطه به سیستم معرفی شد ، یک گزارش کلی از تصاویر موجود گرفته می شود . خروجی این مرحله به شرح جدول ۱ می باشد.

نام گروه تصاویر	تعداد تصویر در هر گروه
تصاویر فرودگاه	۵۵
تصاویر پارک ها	۵۰
تصاویر زمینهای ورزشی	۵۰
تصاویر پارکینگ ها	۵۰

جدول ۱: گزارش اولیه از تصاویر

یادآور می گردد که در جدول ۱ تنها قسمتی از خروجی نمایش داده شده است . پس از ورود تصاویر مربوط به شبیه سازی انجام شده می بایست فرآیند بالانس کردن تعداد تصاویر در هر گروه را انجام داد . در شبیه سازی عملیات بالانس یا متعادل کردن تعداد تصاویر به صورت ساده ای صورت می گیرد . بنابراین خروجی نهایی برای قسمت قبل در جدول ۲ نشان داده شده است. همان طور که در جدول شماره ۲ مشاهده می شود تنها تعداد تصاویر موجود در دسته فرودگاه ها یکسان نبود که این گروه به سمت ۵۰ نمونه بالانس شده اند.

نام گروه تصاویر	تعداد تصویر در هر گروه
تصاویر فرودگاه	۵۰
تصاویر پارک ها	۵۰
تصاویر زمینهای ورزشی	۵۰
تصاویر پارکینگ ها	۵۰

جدول ۲ : گزارش خروجی نهایی تصاویر

۲-۵- پیش آموزش داده ها با استفاده از الگوریتم شبکه عصبی AlexNet Network

الگوریتم شبکه عصبی AlexNet یکی از پرکاربردترین الگوریتم هایی است که امکان آموزش مدل مربوط به روش پیشنهادی را فراهم می سازد . با استفاده از این الگوریتم کلیه تصاویر مورد پردازش قرار گرفته و یک پیش آموزش کلی بر روی آنها صورت می گیرد . الگوریتم شبکه عصبی AlexNet ساختار الگوریتم شبکه عصبی کانولوشن را تعریف می کند که خروجی این قسمت برای تنها تعدادی از نمونه های استفاده شده درجدول ۳ نشان داده شده است.

۱	data'	Image Input	227x227x3 images with 'zerocenter' normalization
۲	conv1'	Convolution	96 11x11x3 convolutions with stride [4 4] and padding [0 0]
۳	relu1'	ReLU	ReLU
۴	norm1'	Cross Channel Normalization	cross channel normalization with 5 channels per element
۵	pool1'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
۶	conv2'	Convolution	256 5x5x48 convolutions with stride [1 1] and padding [2 2]
۷	relu2'	ReLU	ReLU
۸	norm2'	Cross Channel Normalization	cross channel normalization with 5 channels per element
۹	pool2'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
۱۰	conv3'	Convolution	384 3x3x256 convolutions with stride [1 1] and padding [1 1]
۱۱	relu3'	ReLU	ReLU
۱۲	conv4'	Convolution	384 3x3x192 convolutions with stride [1 1] and padding [1 1]
۱۳	relu4'	ReLU	ReLU
۱۴	conv5'	Convolution	256 3x3x192 convolutions with stride [1 1] and padding [1 1]
۱۵	relu5'	ReLU	ReLU
۱۶	pool5'	Max Pooling	3x3 max pooling with stride [2 2] and padding [0 0]
۱۷	fc6'	Fully Connected	4096 fully connected layer
۱۸	relu6'	ReLU	ReLU
۱۹	drop6'	Dropout	50% dropout
۲۰	fc7'	Fully Connected	4096 fully connected layer
۲۱	relu7'	ReLU	ReLU
۲۲	drop7'	Dropout	50% dropout
۲۳	fc8'	Fully Connected	1000 fully connected layer
۲۴	prob'	Softmax	softmax
۲۵	output'	Classification Output	crossentropyex with 'tench', 'goldfish', and 998 other classes

جدول ۳ : 25*1 Layer array with layers

۳-۵- پیش پردازش نمونه ها با استفاده از الگوریتم یادگیری عمیق CNN

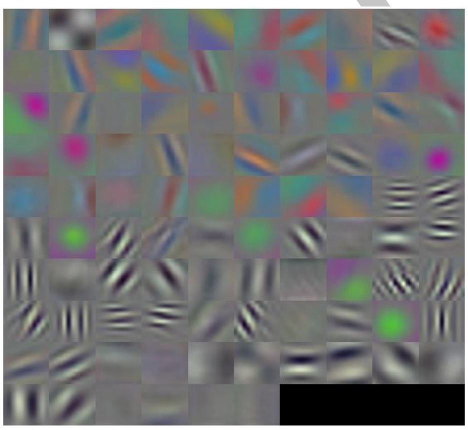
پس از تعیین پیش فرض های لازم و اعمال یک پیش آموزش اولیه با استفاده از الگوریتم شبکه های عصبی کانالوشن لازم است که یک پیش پردازش کلی بر روی نمونه های وارد شده صورت گیرد . بنابراین یک پیش پردازش بر روی کلیه داده ها صورت می گیرد و نمونه های پیش پردازش شده در یک مکان جدا ذخیره سازی و سپس مابقی فرآیند ها بر روی این نمونه ها اعمال می شوند.

۴-۵- آماده سازی داده ها

در این قسمت داده های مربوط به فرآیند آموزش و آزمایش نمونه ها صورت می گیرد . در این پژوهش ۸۰٪ از تصاویر مربوط به هر مجموعه داده به عنوان نمونه های آموزشی جهت تولید مدل های مربوطه تعیین شده و ۲۰٪ از تصاویر مربوط به هر مجموعه داده جهت ارزیابی نتایج روش پیشنهادی تقسیم بندی می شوند.

۵-۵- استخراج ویژگی ها از تصاویر با استفاده از الگوریتم یادگیری عمیق CNN جهت آموزش

در این مرحله به منظور آموزش روش پیشنهادی لازم است ویژگی های بافتی و لبه ای تصاویر استخراج شود . این الگوریتم به صورت مجزا بر روی هر تصویر از هر گروه مجموعه داده ها اعمال شده و ویژگی های هر تصویر را استخراج نموده تا در مراحل بعد، از آن استفاده شود.



شکل ۶ : خروجی لایه اول الگوریتم یادگیری عمیق CNN

در شکل شماره ۶ که خروجی لایه اول از الگوریتم یادگیری عمیق CNN می باشد ، ویژگی های لبه و گرفتن لکه های تصویر نشان داده شده است . بنابراین این فاز جهت استخراج ویژگی ها از تصویر انجام می شود.

۶-۵- اعمال الگوریتم SVM جهت دسته بندی تصاویر

پس از اعمال الگوریتم یادگیری عمیق جهت شناسایی تصاویر توسط CNN ، خروجی مربوط به این فاز به الگوریتم SVM داده شده و در نهایت تصاویر دسته بندی می شوند . پس از اعمال این الگوریتم ، نتایج مورد نظر ارزیابی می شوند که در قسمت بعد مورد بررسی قرار می گیرد.

۷-۵- ارزیابی نتایج روش پیشنهادی

اطلاعات بدست آمده بسیار مبهم می‌آشد زیرا نیاز به تفسیر دارند . در سیستم بازیابی اطلاعات باید هر چه بهتر اطلاعات را مدل کرد تا ابهام در درک اطلاعات توسط سیستم کمتر شوند . به همین علت در سیستم های بازیابی اطلاعات ، معیار دقت و بازخوانی و معیارهایی شبیه به آنها به عنوان معیارهای اصلی ارزیابی به کار می‌روند . به منظور ارزیابی نتایج روش پیشنهادی از معیار های (۱) معیار دقت : تعداد مستندات بازیابی شده واقعاً با ربط باشند یا بعبارت دیگر نزدیک بودن مقدار اندازه گیری به یکدیگر خواه واقعیت را نشان دهد، خواه نشان ندهد (۲) معیار بازخوانی (۳) معیار صحت : نزدیکی مقدار اندازه گیری شده به مقدار واقعی یعنی مقداری که ما به آن اطمینان داریم .

$$\text{Precision} = \frac{TP}{TP + FP} \quad (۳)$$

$$\text{ReCall} = \frac{TP}{TP + FN} \quad (۴)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (۵)$$

جدول خروجی طبقه بندی به کلاس های مثبت و منفی در جدول ۴ نشان داده شده است.

		کلاس تخصیص یافته توسط مدل	
		مثبت	منفی
کلاس واقعی	مثبت	True Positive / (TP)مثبت حقیقی	False Negative / (FN) منفی کاذب
	منفی	False Positive / (FP)مثبت کاذب	True Negative / (TN)منفی حقیقی

جدول ۴ : خروجی طبقه بندی به کلاس های مثبت و منفی

۶- ارزیابی نتایج

آزمایش های ما توسط پروتکل ۵ پوشه Cross-Validation انجام شد که در نتیجه آن ، مجموعه داده ها در پنج پوشه با یک اندازه مرتب شد . به عنوان مثال تصاویر تقریبا به طور مساوی به پنج پوشه بدون تداخل تقسیم شده‌اند. مجموعه داده زمین های قهوه و غیر قهوه دارای ۴ پوشه می باشد که هر کدام ۶۰۰ تصویر دارد و یک پوشه نیز دارای ۴۷۶ تصویر می باشد ، و هر پوشه طوری بالانس شده که هر کدام دارای ۵۰٪ تصویر قهوه و ۵۰٪ تصویر غیر قهوه می‌باشد.

وقتی تنظیم دقیق یا آموزش یک شبکه از صفر اجرا می‌شود ، در هر بار اجرا ، ۳ تا از پوشه ها به عنوان تنظیم آموزش و یکی به عنوان اعتبارسنجی و پوشه باقیمانده به عنوان مجموعه آزمایشی استفاده می‌شود . مهم این است که یادآوری کنیم که وقتی پوشه های train-set و validation وtest را عوض می‌کنیم ، آموزش کامل و تنظیم دقیق شبکه از ابتدا شروع می‌شود به این معنی که برای هر یک از پنج پوشه ، با هر بار انجام مراحل cross-validation ، یک شبکه‌ی متفاوت بدست خواهد آمد.

وقتی از شبکه کانالوشن به عنوان استخراج کننده‌ی ویژگی استفاده می‌کنیم ، چهار مجموعه به عنوان آموزش استفاده می‌شوند درحالی که آخرین آن test-set است . ما معمولا از Linear SVM به عنوان طبقه بندی نهایی استفاده می‌کنیم.

وقتی یک شبکه را تنظیم دقیق کرده یا آموزش کامل می‌دهیم ، پارامتر های نسخه‌ی اصلی را نگه می‌داریم و فقط دو پارامتر را طبق جدول ۵ تغییر می‌دهیم . مهم است که بیان کنیم که وقتی از یک شبکه کانولوشن از پیش آموزش دیده (بدون تنظیم دقیق) به عنوان استخراج کننده‌ی ویژگی استفاده می کنیم ، دیگر هیچ آموزشی نخواهیم داشت از این رو هیچ پارامتری برای تغییر نمی‌ماند .

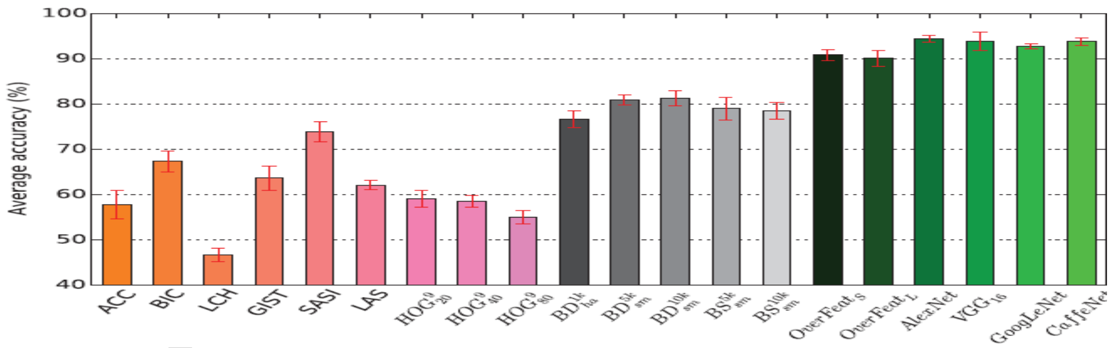
Strategy	#iterations	Learning Rate
Fine-tuning	20,000	0.001
Full-training	50,000	0.01

جدول ۵ : پارامترهای مورد استفاده در استراتژی های تنظیم دقیق و آموزش کامل

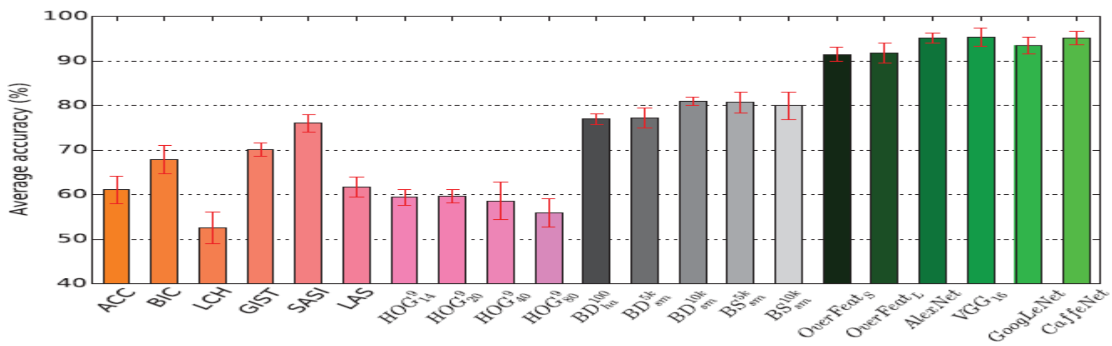
نتایج براساس دقت میانگین و انحراف استاندارد میان پنج پوشه بیان می‌شوند . برای یک پوشه ، ما دقت را برای هر کلاس و سپس دقت میانگین را بین تمام کلاس ها محاسبه می‌کنیم که نهایتا این دقت برای محاسبه‌ی میانگین نهایی در میان پنج پوشه استفاده می‌شود.

تمام آزمایشات بر روی یک سیستم ۶۴ بیت اینتل i5 با فرکانس ۲,۸ هگز و ۸ GB RAM و یک عدد کارت گرافیک GeForce با GB۴ حافظه‌ی داخلی و سیستم عامل سون انجام گردید .

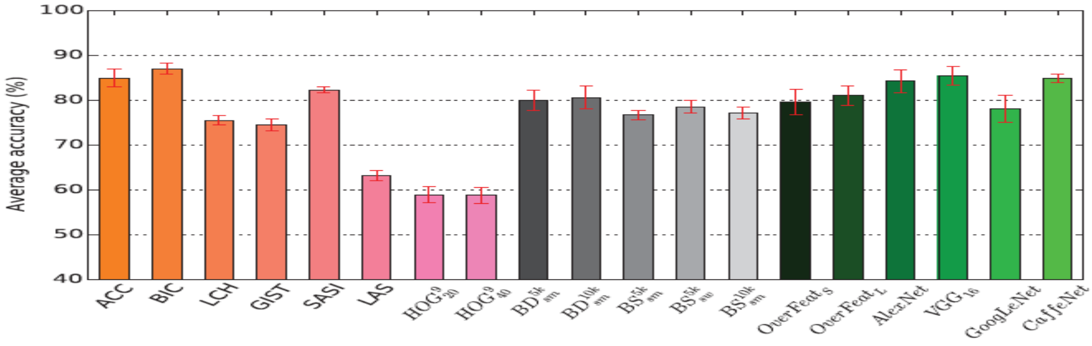
پس از انجام آزمایشات دقت الگوریتم پیشنهادی جهت طبقه بندی تصاویر ثبت شده از راه دور به شرح ذیل است :



شکل ۷: میانگین دقت الگوریتم CNN جهت استخراج ویژگی ها بر روی مجموعه داده UCMerced

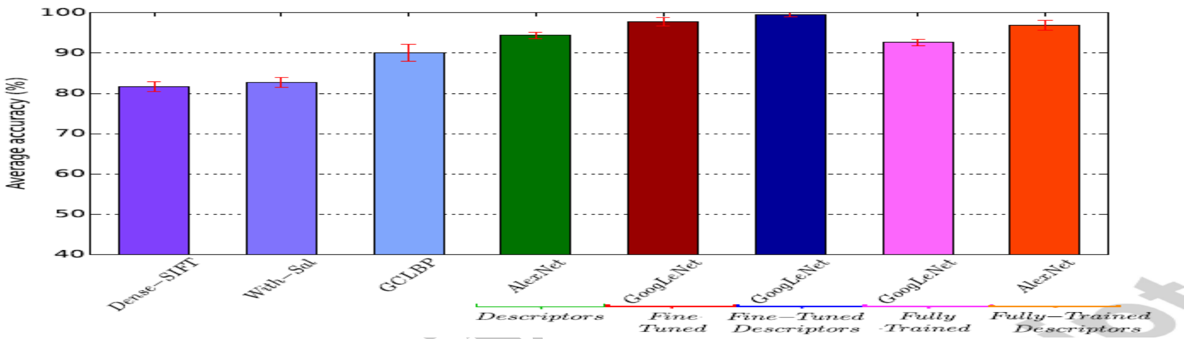


شکل۸ : میانگین دقت الگوریتم CNN جهت استخراج ویژگی ها بر روی مجموعه داده تصاویر گوگل ارث

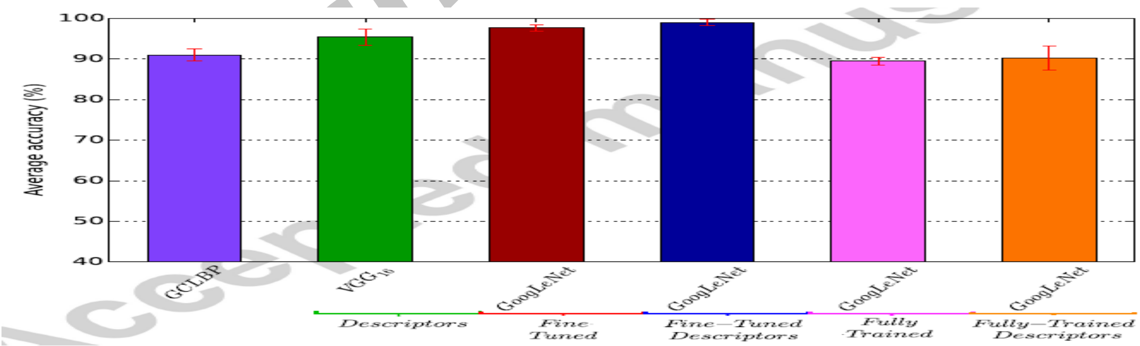


شکل ۹ : میانگین دقت الگوریتم CNN جهت استخراج ویژگی ها بر روی مجموعه داده زمین های قهوه

یادآور می گردد که در شکل های ۷ تا ۹ دقت استخراج ویژگی های مختلف از تصاویر توسط الگوریتم شبکه عصبی کانولوشن نشان داده شده است . همچنین در شکل های ۱۰ تا ۱۲ نیز میزان دقت انواع الگوریتم های یادگیری عمیق جهت استخراج

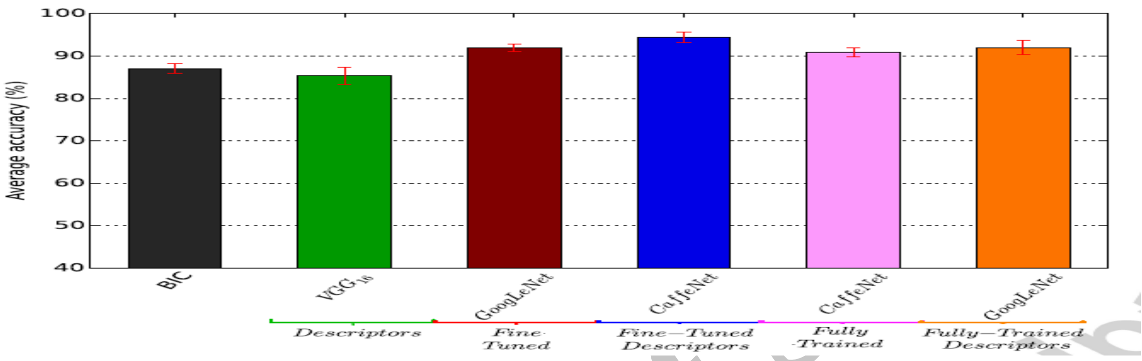


ویژگی های تصاویر ثبت شده ازراه دور نشان داده شده است .



شکل۱۰ : میانگین دقت الگوریتم های یادگیری عمیق جهت استخراج ویژگی ها بر روی مجموعه داده UCMerced

شکل ۱۱ : میانگین دقت الگوریتم های یادگیری عمیق جهت استخراج ویژگی ها بر روی مجموعه داده تصاویر گوگل ارث



شکل ۱۲ : میانگین دقت الگوریتم های یادگیری عمیق جهت استخراج ویژگی ها بر روی مجموعه داده زمین های قهوه

با توجه به مطالب عنوان شده می توان نتیجه گرفت که هر کدام از الگوریتم های یادگیری عمیق دارای دقت مختلفی بر روی همه داده های استفاده شده هستند .

۷- نتیجه گیری

جهت نوآوری و جدید بودن تحقیق، با بررسی سوابق پیشین و پژوهش هایی که در زمینه طبقه بندی تصاویر راه دور انجام شده است ، مشاهده گردید که نوع داده ها تا حدودی در دقت استخراج ویژگی ها از تصاویر از راه دور موثر هستند . همچنین مشخص گردید که استفاده از الگوریتم های طبقه بندی مختلف مانند الگوریتم های ماشین بردار پشتیبان ، شبکه های عصبی و درخت تصمیم و غیره جهت طبقه بندی تصاویر دارای دقت های مختلفی هستند . طبق مطالعات صورت گرفته مشخص گردید تاکنون از ترکیب الگوریتم های شبکه عصبی و درخت تصمیم در قالب یک سیستم جمعی و الگوریتم یادگیری عمیق شبکه های عصبی کانالوشن استفاده نشده است که می تواند دقت مناسب تری نسبت به الگوریتم ماشین بردار پشتیبان داشته باشد از این رو ، ترکیب تکنیک های ذکر شده با هم به منظور طبقه بندی تصاویر ماهواره ای از مهمترین یافته های این تحقیق می باشد.

منابع

[1] G. Kumar, P. K. Bhatia, A detailed review of feature extraction in image processing systems, in: Advanced Computing & Communication Technologies, IEEE, 2014, pp. 5{12.
[2] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436{444.
[3] M. Faraji, J. Shanbehzadeh, Bag-of-visual-words, its detectors and descriptors; a survey in detail, Advances in Computer Science: an International Journal 4 (2) (2015) 8{20.
[4] C.-F. Tsai, Bag-of-words representation in image annotation: A review, ISRN Arti_cial Intelligence 2012.
[5] J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, J.-M. Geusebroek, Visual word ambiguity, Transactions on Pattern Analysis and Machine Intelligence 32 (2010) 1271{1283.
[6] K. E. A. van de Sande, T. Gevers, C. G. M. Snoek, Evaluating color descriptors for object and scene recognition, Transactions on Pattern Analysis and Machine Intelligence 32 (9) (2010) 1582{1596.
[7] J. Sivic, A. Zisserman, Video google: a text retrieval approach to object matching in videos, in: International Conference on Computer Vision, Vol. 2, 2003, pp. 1470{1477.
[8] F. Perronnin, J. S_anchez, T. Mensink, Improving the Fisher Kernel for Large-Scale Image Classi_cation, in: European Conference on Computer Vision, 2010, pp. 143{156.
[9] Y.-L. Boureau, F. Bach, Y. LeCun, J. Ponce, Learning mid-level features for recognition, in: Computer Vision and Pattern Recognition, 2010, pp. 2559{2566.
[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248{255.
[11] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Computer Vision and Pattern Recognition, IEEE, 2014, pp. 580{587.
[12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep learning, book in preparation for MIT Press (2016).
URL http://goodfeli.github.io/dlbook/
[13] Y. Bengio, Learning deep architectures for ai, Foundations and trends in Machine Learning 2 (1) (2009) 1{127

[14] O. A. Penatti, K. Nogueira, J. A. dos Santos, Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?, in: Computer Vision and Pattern Recognition Workshop, IEEE, 2015.
[15] K. Nogueira, W. O. Miranda, J. A. Dos Santos, Improving spatial feature representation from aerial scenes by using convolutional networks, in: Graphics, Patterns and Images (SIBGRAPI), 2015 28th SIBGRAPI Conference on, IEEE, 2015, pp. 289{296.
[16] J. Yue, W. Zhao, S. Mao, H. Liu, Spectral{spatial classi_cation of hyperspectral images using deep convolutional neural networks, Remote Sensing Letters 6 (6) (2015) 468{477.
[17] O. A. B. Penatti, E. Valle, R. da S. Torres, Comparative study of global color and texture descriptors for web image retrieval, Journal of Visual Communication and Image Representation 23 (2) (2012) 359{380.
[18] M. Xie, N. Jean, M. Burke, D. Lobell, S. Ermon, Transfer learning from deep features for remote sensing and poverty mapping, arXiv preprint arXiv:1510.00098.
[19] K. Chat_eld, K. Simonyan, A. Vedaldi, A. Zisserman, Return of the devil in the details: Delving deep into convolutional nets, arXiv preprint arXiv:1405.3531.
[20] F. Hu, G.-S. Xia, J. Hu, L. Zhang, Transferring deep convolutional neural networks for the scene classi_cation of high-resolution remote sensing imagery, Remote Sensing 7 (11) (2015) 14680{ 14707.
[21] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.
[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, arXiv preprint arXiv:1409.4842.
[23] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classi_cation with deep convolutional neural networks, in: Neural Information Processing Systems, 2012, pp. 1106{1114.